

Claremont Colleges Scholarship @ Claremont

CMC Senior Theses

CMC Student Scholarship

2015

It's Worse Than We Think: Why It Matters That We Underestimate Depression

Tess Hubbeling

Claremont McKenna College

Recommended Citation

Hubbeling, Tess, "It's Worse Than We Think: Why It Matters That We Underestimate Depression" (2015). *CMC Senior Theses*. Paper 1001.

http://scholarship.claremont.edu/cmc_theses/1001

This Open Access Senior Thesis is brought to you by Scholarship@Claremont. It has been accepted for inclusion in this collection by an authorized administrator. For more information, please contact scholarship@cuc.claremont.edu.

CLAREMONT MCKENNA COLLEGE

**It's Worse Than We Think:
Why It Matters That We Underestimate Depression**

SUBMITTED TO

PROFESSOR S. ANDREW SCHROEDER

AND

DEAN NICHOLAS WARNER

BY

TESS ALON HUBBELING

FOR

SENIOR THESIS

FALL 2014

DECEMBER 1, 2014

Table of Contents

<i>Acknowledgements</i>	<i>ii</i>
<i>Abstract</i>	<i>iii</i>
Introduction	1
Chapter I: Health States and Quality Weights	6
1. How We Compare Health States That Seem Incomparable.....	6
2. Quality Weights – Preference-Based Elicitation Methods.....	9
a. Visual Analogue Scale (VAS)	
b. Standard Gamble (SG)	
c. Time Trade-Off (TTO)	
d. Person Trade-Off (PTO)	
e. An important Caveat	
3. Cost-Effectiveness Analysis and Quality Weights.....	13
Chapter II: The Disability Paradox	16
1. Patient Responses.....	17
2. Non-Patient Responses.....	20
3. Patients vs. Non-Patients.....	26
a. Ethical and Economic Arguments for Choosing Patient or Non-Patient Evaluations	
Chapter III: The Anomaly of Depression	34
Chapter IV: Reframing the Question	39
1. Higher vs. Lower Quality Weights.....	39
Chapter V: Economic Trade-Offs	43
Conclusion	47
Bibliography	50
Appendix	59

Acknowledgements

First and foremost, I need to thank my advisor, Professor Andrew Schroeder, for his endless patience and support throughout the process of writing my thesis. His assistance in shaping my argument as well as finding and vetting relevant literature would have alone been cause for enormous amounts of gratitude. But Professor Schroeder consistently went above and beyond to help me with this project, answering my frantic, late-night emails as the deadline drew near, keeping me on topic when I threatened to veer off on irrelevant philosophical tangents, and supporting me every step of the way. Our weekly thesis meetings were as valuable as any class – each time I left inspired, interested, and incredibly humbled. As in every class I have been lucky enough to take with him thus far, I will be incredibly sorry to see it end.

I must also thank the legions of peers and friends who tolerated me discussing my argument endlessly with them, and either tried to help and humor me by providing counterarguments and insights when they could, or push me by pointing out the gaping holes in my increasingly preposterous analogies. I owe many thanks to Rita Giles, Talia Segal, Charlie Montgomery, Gracie Coburn, Dana Gibson, Jeff Hochhauser, Clay Spence, Maddy Stein, Louisa Dunwiddie, Aviv Caspi, Michael Irvine, Maya Sandalow, Harry Pellicoro, Chase Pribble, Allie Ruark, Remy Pinson, the RAs, the entire Poppa Hall of Fame (and Wall of Shame) and many others who did me the favor of at least pretending to look interested whenever I felt the need to wax loquacious about my thesis.

I finally must thank my mother, father, and sister, who valiantly read through 40 pages of amateur philosophical meandering, which they promised was fun and interesting but I understand was probably more to support me. In the face of my self-doubt they bolstered my confidence at every turn. Their encouragement and support kept me afloat, and I am not confident I could have finished my thesis without them.

Abstract

This paper will examine specific processes involved within the decision-making process of how to allocate limited health care resources. I will start by discussing how in order to compare and differentiate between health states, we have created ranking systems, based on the health state's impact on people's quality of life, which health states need more care, and which can be most effectively treated. We evaluate impact on quality of life by assigning quality weights to years of life lived with that health state, which we call quality-adjusted life years, or QALYs.

Next, I will discuss the problems with assigning quality weights to health states; specifically, the disability paradox, meaning the distinct differences between quality weights assigned by non-patients versus patients.

After that, I will explain how depression defies the trend of the disability paradox, and causes our prior arguments about why patients and non-patients rate health states different to contradict themselves., This leads me to suggest that we should consider a different way of deciding between different quality weights. I examine the arguments for choosing higher or lower quality weights, and conclude that because we have a moral imperative to provide health care resources to those in need, particularly those who are disadvantaged, we should take the lower quality weights and err on the side of overspending on health states. Ultimately, this will create the greatest change in funding for health states like depression that go against the disability paradox. Finally, I address the economic trade-offs we have to consider if we make the decision to spend more on treating health states.

INTRODUCTION

Malaria, heart disease, cancer, depression – all of these diseases kill many people each year. Ideally, we would have unlimited resources so we could fund research teams to hunt for cures for these diseases, and give doctors leeway to administer any and all available treatments to those suffering from these diseases and in need. But resources are limited, and we can't serve every need. Should one of these get priority over the others, or does one merit less attention and funding? How do we determine where and how to use our limited resources?

Consider the difficulty in this hypothetical choice; we have the money to cure one group of people, and we must choose between curing 100 people of depression or 100 people of paraplegia.¹ If our goal is to maximize social welfare, which one of these is more valuable in serving that goal? We need a method of comparing the two, but they lack factors in common, like symptoms or causes, that would allow us to place them on the same scale and make a rational decision that goes beyond some loose intuition of which is “worse.” Deciding how to allocate health care resources will involve making decisions between treating and funding research into health states that have little to

¹ This is probably an unrealistically stark and straightforward scenario, but the real trade-offs would likely be just as harsh and the choices just as difficult to make.

nothing in common; we need a way of differentiating and drawing comparisons between broken limbs and Parkinson's and melanoma.

In order to find the most socially beneficial way of distributing health care resources, we create methods of ranking health states according to their impact on people's lives, we examine which health states need more care, and we evaluate which can be most effectively treated. The primary way that we have chosen to do this is to examine the impact of a health state on the quality of life one can experience. Quality of life encapsulates many factors, which means it can lead to a scale on which we can compare any two health states – any given health state will have some impact on quality of life, even if it is arguably negligible, and thus we can place any health state somewhere on the scale between death and perfect health.²

To measure how health states impact quality of life, we determine the quality weight of a year in a given health state. The weights are on a scale of 0, meaning death, to 1, indicating perfect health. If a person's quality weight with tuberculosis is equal to .669, this would mean that each year of their life was equivalent to roughly 67% of a year of life at full health (Salomon *et al*, 2010). Years described by quality weights are called quality-adjusted life years, or QALYs.

We begin calculating quality weights and thus QALYs by giving surveys, designed by experts, to the general public. These surveys elicit the preferences of the population in regards to different health states and are then converted into quality

² The point at which a change in quality of life becomes negligible is not discussed in this paper, but the larger point, that every health state other than perfect health will have some impact, is important to note in order to show why quality of life is a good metric to use.

weights, also called preference scores, for different health states. We use these quality weights in cost-effectiveness analysis to compare the gain from treating or curing different health states. For example, if someone has a QALY per year of .4, curing them would incur a QALY gain of .6. On the other hand, a person with a different disease might have a QALY of .2 and curing them would incur a QALY gain of .8. If costs were equal we would put our limited health care resources into helping the person with more potential for gain. Even in scenarios when costs are equal, we can calculate cost-effectiveness ratios in order to compare and make choices.

This would all be very straightforward if we had a single, precise value for the quality weight of each health state. However, this is not the case. Different groups (i.e. different age brackets, genders, etc.) sometimes seem to provide different quality weights for the same health state. One of the most studied and most debated of these is the differences in the average quality weights given by patients versus non-patients.

Patients who have a specific health state and non-patients who have never experienced it tend to give very different numbers for the same health states. This leads to many arguments for which group we should listen to when we use quality weights in cost-effectiveness analysis, as the different numbers lead to different cost-effectiveness ratios. Many authors have given different reasons for the variation between these groups and what factors may lead one group to have an inaccurate understanding of their own quality of life.

This becomes even more complicated when we consider the case of depression. Normally, patients give higher quality weights to health states while non-patients give

lower ones. This is called the Disability Paradox, and will be discussed at length in chapter two, The Disability Paradox. But evaluations of depression show the opposite trend, with patients valuing depression at a lower quality weight than non-patients. The implications of this might lead us to consider taking an alternate approach in trying to determine which quality weight to take instead of trying to choose between patients and non-patients. Instead of trying to see which group, patients or non-patients, is right, perhaps we should consider the possibility that neither group is wrong. I will later discuss the idea that both groups are merely evaluating the health states from different, but both accurate, perspectives. If this is the case, deciding between patients and non-patients is not the right choice to make. Instead, we might consider the problem from a more abstract standpoint, considering instead the ethical and economic implications of choosing either a higher or a lower quality weight, regardless of which group provided that number. Ultimately, I posit that we should take the lower number. In many instances this will imply more spending on health states; however, I believe it is our moral imperative to give more rather than less to help treat health problems.

Before we can reach this conclusion and understand why depression causes us to consider the question of which quality weight to choose differently, we need to understand both the context of the question (in terms of deciding how to allocate scarce health care resources) and also the other argument between patient and non-patient assigned quality weights. I start with the background. First, I will explain our different methods for finding quality weights and how they are used in allocating health care resources. Next, I will discuss the disability paradox, the problems it poses, and the arguments people give for either patients or non-patients. After that, I will explain how

depression defies the trend of the disability paradox, and causes our prior arguments to contradict themselves, thus leading us to a different understanding of how to decide different quality weights. I examine the arguments for choosing higher or lower quality weights, and conclude that because we have a moral imperative to provide health care resources to those in need, particularly those who are disadvantaged, we should take the lower quality weights and err on the side of overspending on health states. Ultimately, this will create the greatest change in funding for health states, such as depression, that go against the disability paradox. Finally, I address the economic trade-offs we have to consider if we make the decision to overspend.

CHAPTER I: HEALTH STATES AND QUALITY WEIGHTS

How We Compare Health States that Seem Incomparable

When evaluating the effectiveness of different healthcare treatments, we consider several factors: how the treatment will affect life expectancy (mortality) and how it will affect the quality of life experienced during those years (morbidity) (Gudex *et al.* 1996, Weinstein, Torrance, and McGuire, 2009). Measuring just one side of this equation will not adequately capture how a treatment can impact someone's life. We can see this if we consider the following example.

Two people are in a car together and get into a bad accident. One ends up paralyzed below the waist and severely mentally handicapped due to head trauma. The other breaks their leg. Let's assume that neither person's life is shortened due to the accident and ensuing injuries. If we only measure impact on mortality, both people are alive, and therefore the "same." But this seems wrong for several reasons; our intuition alone tells us that a broken leg is not the same as paraplegia and a mental handicap. We can also see how the person with paraplegia will need more care and medical attention than the person with the broken leg in order to function in daily activities. Thus, we can see the value in

measuring both quality and quantity of life when evaluating the effectiveness of treatments.

This makes it clear that in order to allocate healthcare effectively, researchers need a measure that captures the impact of a health state on both mortality and morbidity. In order to do this, we assign a value to the quality of life one experiences each year as a result of a health state, thus weighting each year according to its quality (Nord, 1992, Guyatt, 1993, Brazier, 1999, Weinstein, Torrance, and McGuire, 2009). This will reflect the impact on length of life as well as the change in quality. We call this measure quality-adjusted life years, or QALYs. Health economists can use changes in QALYs to assess the benefits of treatments; the degree to which a treatment improves quality or length of life will be captured by a QALY measure.

In order to calculate QALYs we first need to assign quality weights to health states.³ We could have health experts assign different values to each health state according to their expert assessment. But we tend to believe that it is valuable to incorporate not just the judgments of health care professionals, but also the preferences of the general population, when attempting to assign numbers to the quality of life experienced with different health states (Nord, 1992, Schwappach, 2002). Since QALYs are used on a societal and not just an individual level, it seems valuable to have a number that reflects society's values, not just expert ones. Furthermore, all people have a vested interest – not only is the money for health care spending coming predominately from tax dollars, people are all potential patients for some if not most health state (Nord, 1992, Gudex *et al.* 1996).

³ Quality weights can also be called health state utilities.

So in the interest of incorporating the public's preferences for different health states, we conduct special types of surveys, designed by experts to elicit people's judgments of the quality of life attached to various health states. These surveys are sometimes called preference elicitation surveys, and the responses represent the quality weights people assign to different health states. There are four main preference elicitation surveys. These are visual analogue scale (VAS, or Rating Scale, RS), standard gamble (SG), time trade-off (TTO), and person trade-off (PTO) (Brazier, 1999).

While each of these four approaches will have a different metric, all responses can be converted into quality weights. The scale for quality weights ranges from 0, representing death, to 1, representing full or perfect health (Nord, 1992, Gudex *et al.* 1996, Brazier, 1999). A quality weight of .95 would indicate a health state with only a small impact on quality of life, while a quality weight of .4 for one life year would indicate very poor quality of life.

Changes in quality weights or QALYs are predominately used in cost-effectiveness analyses (Weinstein, Torrance, and McGuire, 2009, Broome, 2003). The impact of a treatment for a health state can be measured in the relative gain or loss of quality-adjusted life years, and researchers or analysts can compare relative gains or losses to costs to evaluate cost-effectiveness (Weinstein, Torrance, and McGuire, 2009, Broome, 2003). Quality weights and QALY values can help analysts rank health states in terms of impact on quality of life – they will help us see clearly which health states are the most and least severe, and thus, from a governmental point of view, will guide us in determining where to allocate money for research into more effective treatments, and what to cover with

insurance or national health care (Weinstein, Torrance, and McGuire, 2009, Broome, 2003).

This chapter will briefly explain how each of these preference-based health state valuation methods works before explaining the other role that health state utilities play, as metrics of effect in cost-effectiveness analyses.

Quality Weights – Preference-Based Elicitation Methods

As mentioned above, there are four methods used to elicit people's judgments of the quality of life attached to various health states. All of the methods of preference-based elicitation have the same goal of applying a quantitative value to the quality of life experienced by a patient with a particular health state. They all attempt to elicit the value an individual would place on being in or escaping a particular health state. Each approach asks a slightly different question (Brazier, 1996, Bleichrodt and Johannesson, 1997).

Visual Analogue Scale, (VAS or Rating Scale, RS)

This method is the most straightforward of the four approaches. The rating scale method asks respondents to place several different health conditions on a scale to directly rank them against each other. This is often done by literally giving the patients a scale on a piece of paper and asking them to place the health states in question on the line. The ends of the line are typically death or worst possible health on one side, and perfect

health on the other. Respondents may be asked to rank many different health states at once or place them on the scale one at a time. The line can either be vertical or horizontal, can range from 0 to 1 or from 0 to 100, and may or may not have marked intervals (Brazier *et al.* 1999). Rating scale is widely used to evaluate health states, in part because it simple and easy to use (Brazier *et al.* 1999).

Standard Gamble (SG)

Standard Gamble questions ask people about their willingness to risk their life or health in order to escape a particular health state. A standard gamble question would ask: “if you could undergo a treatment that would cure you of Health State X, what risk of death due to that treatment would you accept and yet still choose to undergo the treatment?” To use a concrete example, a standard gamble question might ask the respondent to give the risk of death they would accept in order to be cured of blindness. The answers are given in a percentage scale; an answer closer to 100% would indicate that a respondent considers that health state particularly terrible. If you are willing to accept a 60% chance of death to be cured of a health state, you clearly consider it to have a very low quality of life. Whereas if you are only willing to accept a 2% chance of death for a cure to a particular health state, you probably think that health state isn’t all that bad, and you would rather not risk death to cure yourself of it (Green *et al.*, 2000, Brazier *et al.* 1999).

Standard Gamble questions were developed based on Von Neumann and Morgenstern’s Expected Utility Theory (Green *et al.* 2000).

Time Trade-Off (TTO)

The time trade-off method is closely related to standard gamble. Torrance *et al.* (1972) developed TTO as a counter to standard gamble (Brazier *et al* 1999). Instead of asking people to consider what risk of death they would accept, it asks respondents to determine a number of health years they would be willing to give up in order to be cured of a certain health state (it is assumed that the health state has no impact on life expectancy and that one would live an average number of years unless one chose to make the trade). A time trade-off question would ask: “Consider that you have Health State X, and that you have 65 years left to live. How many years of life would you give up to live in full health?” So in response to a time trade-off question, someone might say that they would give up 15 years of life with depression to live in full health, meaning that they would rather live for 50 years in full health over 65 years with depression. The more years of life given up, the more severe and terrible an individual considers a health state (Green *et al.* 2000).

Person Trade-Off (PTO)

The first three methods ask the respondent to consider themselves and what value they personally place on different health states. Person trade-off questions, on the other hand, ask respondents to consider how they would evaluate others with health states with themselves as an impartial third party.

An individual is asked to compare two different health outcomes to indirectly state which is worse than the other. A person trade-off question would ask: “imagine you can cure 100 people of Disease A, or you can cure X number of people with Disease B. What number x would there have to be to cause you to choose to cure that group of Disease B?” If you choose a larger number than 100 to cure of Disease B, you consider Disease A worse. If you choose a smaller number than 100, you consider Disease B worse. For example, someone might say that they would cure 70 paraplegics over 100 blind people because they consider paraplegia relatively worse than blindness, to that particular degree. But alternatively, in the choice between 100 blind people and 60 paraplegics, they might make the opposite choice, and choose the blind patients instead, depending on the relative weight they place on each health state (Green *et al.* 2000).

An important caveat:

Much literature has been devoted to the fact that each of these four methods, SG, TTO, PTO, RS, have the problematic tendency of providing different quality weights for the same diseases (Brazier *et al.* 1999, Essink-Bot *et al.* 1990, Green, Brazier, & Deverill, 2000, CITE?). For example, take the case of a moderate physical disability described by Bleichrodt and Johanneson (1996). Respondents rated the health state using RS, TTO, and SG. Rating scale gave a quality weight of .406, time trade-off gave a value of .578, and SG gave a value of .671. These answers are significantly different at a 99% confidence level (Bleichrodt & Johanneson, 1997).

It is important to note that we will probably never get a unified answer, as each question is asking something different. However, it is worth considering that we may need to decide which elicitation method is most valid and predominately use that particular approach. While the question of which survey type to use and which is more accurate merits plenty of attention, it is not the focus of my argument, and thus I set aside this topic in order to consider other questions.

Cost-Effectiveness Analysis and Quality Weights

Quality weights play a dual role. They can be used for ranking health states as previously discussed, but they also matter in terms of evaluating the effectiveness of treatments. In cost-effectiveness analysis, we can compare the net QALY gain in order to evaluate the benefit of a particular health intervention, and then we can compare that against the cost of the treatment in order to determine a cost-effectiveness ratio.

Because QALYs play a unique role as a metric by which we can compare wildly different health states, their incorporation into cost-effectiveness analysis lends it a similar universality. We can use cost-effectiveness analysis of cost per QALY gained to compare two different treatments aiming to improve the condition of someone with the same disease, or compare between two treatments for different diseases.

Consider the following hypothetical example. You are a patient with arthritis. Treatment A will give you .15 QALY gain for a cost of \$600. Treatment B will give you a QALY gain of .2 for \$1000. If you calculate the ratio of cost of treatment to the QALY gain, you can see that Treatment A is the more efficient option, since it is lower cost per QALY gained, and thus even though your QALY gain is slightly smaller, it produces a better ratio of cost to effect.

We can also use comparisons of cost per QALY gained to compare treatments that are targeting different health states. Imagine the following scenario. You are choosing between two patients. One has terrible migraines, which has a quality weight of .567 (Salomon *et al.* 2010). The other has untreated and uncontrolled asthma, with a quality weight of .868 (Salomon *et al.*, 2010). The money you have, \$300, dollars, is exactly the right amount to cure either of them, returning either to a QALY per year of 1, or perfect health.

The QALY gain from curing the person with migraines is .433 per year, while the gain from curing the person with asthmas is .132 QALYs per year (Salomon *et al.*, 2010). The treatment for the patient with migraines is clearly most cost-effective because it produces a better return for the same dollar amount, and thus the cost-effective analysis would suggest choosing to treat migraines over asthma in this scenario.

These examples show the role that quality weights can play in determining the cost-effectiveness of different treatments. This would be very straightforward analysis if we had a definitive quality weight for all health states. However, researchers realized that different respondent groups might have different biases causing them to respond in different ways.

For example, it seems plausible to wonder if age would provide different responses to something like time trade-off, or if certain levels of education would make one more or less risk-averse. A study done by Sackett and Torrance found some evidence suggesting that age had impacts on the way respondents evaluated health states (1978). Ubel *et al.* found some similar evidence that age had an impact on the way that people evaluated health states (2005). However, Salomon *et al.* in the Global Burden of Disease Study in 2010 found very little evidence of demographic context altering results after doing large-scale comparisons of results from extremely diverse groups (2010).

However, there is one divide that seems to fairly clearly lead respondents to give responses that are significantly different; patients, when evaluating a health state that they have experienced or are experiencing, give answers that are different from those given by non-patients who have never experienced that same health state (Albrecht and Devlieger, 1999). This is often called the Disability Paradox.

CHAPTER II: THE DISABILITY PARADOX

If we look at the health state evaluations provided by patients and non-patients separately instead of aggregately, we discover an apparent paradox, called the Disability Paradox (Albrecht and Devlieger, 1999). The popular conception is that to have high quality of life, one must have full health – but patients' reported quality of life defies this logic (Goode, 1994, Albrecht and Devlieger, 1999). Patients consistently rate their own health states as higher than the general public rates them (Albrecht and Devlieger, 1999).

Consider kidney disease being treated with dialysis. Non-patients who have never undergone dialysis treatments value life with dialysis at .39, which means they would agree to a treatment with a greater than 60% risk of death if there was a 39% chance it would rid them of kidney disease. Another way to consider it is that they would give up more than 60% of their remaining lifespan in order to enjoy full health and not have to undergo dialysis. In contrast, dialysis patients placed their quality of life with dialysis at .56 (Sackett and Torrance, 1978, De Wit, Busschback, and De Charro, 2000). This means that patients would give up just barely more than 40% of their remaining life years to be free from dialysis, a percent dwarfed by the 60% that non-patients are willing to forego.

The Disability Paradox is thus a paradox because the patients' ratings of their quality of life in various health states do not match the expectations the rest of the population holds for what their quality of life would be. As Albrecht and Devlieger point

out, a high quality of life is normally characterized as one with full health, complete life satisfaction, and an overall positive experience of well-being (Albrecht & Devlieger 1999). When we consider someone with a severe and chronic disability (disability is the term employed by Albrecht and Devlieger and thus repeated here), by definition they lack “full health.” Hence our confusion and the paradox; how can someone without access to full health experience a state we traditionally define as including full health?

The Disability Paradox becomes problematic when we consider the role of quality weights in cost-effectiveness analysis. We no longer have one quality weight assigned to a health state, we have two. Thus we are forced to consider the question; who should we listen to, patients or non-patients? The different quality weights they provide will lead to very different cost-effectiveness ratios. We need a singular evaluation of health state utility, so we need to decide which quality weight to consider as the value for that particular health state.

There are arguments on both side of the aisle. But before we examine those and the pros and cons of picking one group over the other, let us determine if there are any factors involved in the survey process or any features of either group that might suggest that one group is “wrong.”

Patient Responses

First, let’s look at patients, and attempt to determine if there is anything that might suggest some sort of error. We can see if there is any factor about patients could cause

their assessments of the quality of life they experience with a health state to no longer serve as an accurate reflection.

One possibility is that patients might alter their scale of quality of life in response to a new health state; in other words, their “10” with a disease would be equivalent to a “7” on the scale of a healthy person. Quality scales are open to personal interpretation and subjectivity, and it seems plausible that personal assessments of quality could change in relation to life events. Individual definitions of “high quality of life” have the potential to vary wildly, particularly if one’s personal frame of reference has been irrevocably shifted thanks to a disability or a disease (Ubel, Loewenstein, & Jepson 2003). This phenomenon is referred to as “scale recalibration” or “response shift” (Ubel *et al.* 2005, Ubel, Loewenstein, & Jepson 2003, Schwartz *et al.* 2006).

In general, increasing sample size could control for this potential for error by generalizing towards a mean. But if this is not the case, we should consider whether or not patients of a particular health state might have unilaterally recalibrated their sense of high quality of life in accordance with their new health state.

This is reinforced by the fact that research has shown that the conversational context of a survey can shift a patient’s response, indicating that their scale of quality of life is potentially largely (and at least partially) relative. This was tested by conducting interviews in which the interviewers gave different introductions to different groups of respondents participating in phone interviews (Ubel *et al.* 2005). One version of the introduction highlighted the interviewee as a patient, noting their status as a member of a group with a particular health state, and asked about their quality of life relative to that

group. The other version of the survey introduction cited the individual as part of a larger group unaffiliated with that health status.

The results of this test showed that conversational context could indeed influence an individual's reported quality of life. Individuals who compared their quality of life with others who shared their health state rated their quality of life as on average higher than those who compared it to a population who did not have their health state. The researchers argued that this showed that patients were considering that on a scale where everyone was in their health state, they were relatively high, but then their answers shifted downwards when they compared themselves to "normal" health (Ubel *et al.* 2005).

Ubel *et al.* also conducted a study to test for overall scale recalibration (2005). Respondents were given slightly varied interview questions depending on different demographic factors, and it showed that people have some tendencies to adjust their personal health scale based on age and health state. However, further studies conducted by Ubel and others have sought to show the impact of scale recalibration and found very little evidence to support its occurrence. This suggests that scale recalibration, even in those small incidences where it occurs to a small degree, is not an important factor in driving the disability paradox. These conflicting results mean that we cannot be sure whether or not scale recalibration causes patients to rate their quality of life as higher than they should (Schwartz *et al* 2006).

Non-Patient Responses

Now let's turn to non-patient responses, and see if there is something about the way that the general population evaluates quality of life that could indicate that the quality weights they assign to health states are too low, and do not accurately reflect the quality of life experienced in different health states.

There is a fairly large body of evidence that suggests that people are bad at predictions of future states, particularly hedonic predictions. This is commonly attributed to several factors, two of which could play large roles in causing non-patients to assign lower quality weights to health states than they actually merit. (There are many other factors possibly at fault, but for now I am addressing only these two major potential causes in detail.)

The first potential factor is a focusing illusion. A focusing illusion refers to when we overemphasize the things that will change thanks to an event and fail to note what will stay the same (Ubel, Loewenstein, & Jepson 2005). Consider a case where you are in a bad accident and end up having to amputate both legs below the knee. When people picture the impact that will have on their lives and how they will feel after the event, they consider limited mobility and the ways in which that will change their current behaviors; they worry about not being able to walk or run again.

Fixated on the negative impacts, people fail to note the multitude of positive aspects in their lives that would be unaffected by the accident. They could still have a happy life with a loving family, play with pets and friends and children, listen to music, watch sunsets, enjoy great food and company, and more; essentially, a person could

engage in nearly every activity that someone might associate with a wonderful life. While their quality of life might decrease somewhat relative to their life before the accident (though even that is up for debate in cases such as deafness) it is very possible that it is not as bad as they think it will be. This could indicate that non-patients underestimate the quality of life that those living with a health state experience.

Failing to consider or adequately predict adaptation is another potential factor (Ubel, Loewenstein, & Jepson 2003). People often fail to perceive how well they will be able to emotionally adapt to new circumstances, which they are in fact quite good at doing (Ubel, Ubel *et al* 2005, Ubel, Loewenstein, and Jepson, 2003). Emotional adaptation can occur consciously, when people work to find new means of emotional fulfillment or attaining happiness (Ubel, Loewenstein, & Jepson, 2003). For example, in the case of the amputee, they may have originally loved dancing, but now realize and develop a deep love for music, from which they gain the same level of joy as they gained from dancing.

It can also happen unconsciously; studies show that over time, both positive and negative emotional reactions fade. When predicting how an event or life change will make them feel, people can rarely see past the initial, strong emotional reaction, and they fail to consider how that will dissipate over time.

There are several different ways in which we fail to foresee how we will emotionally adapt, and this means that non-patients will tend to underestimate how well they will be able to cope with the loss of their legs. You might only consider one thing that you will be unable to do such as a particular hobby you can no longer pursue, and fail to appreciate the broader context of other activities that share similar value and

function. In the case of the amputee, you might be an avid runner who likes to race, and saddened by the fact that you will no longer be able to do this, thus missing the fact that you could enjoy a similar type of physical competition by participating in wheelchair races. Or, you might be so intent upon how one thing will negatively impact you, and fail to see other more positive events that will occur and balance it out. For example, you might be solely thinking about the impact on your mobility throughout your lifetime, and fail to consider the many positive events that will occur at the same time, like falling in love, starting a family, and more. Finally, people tend to have trouble remembering both the degree to which and how quickly emotional reactions fade. While you might be initially very upset by the loss of your legs, humans are adept at pushing away negative emotions, and thus the sharpness of your emotional pain will fade more rapidly than you tend to expect.

Before we accept that non-patients are inaccurately evaluating quality of life for health states, we need to consider several factors that play a role in whether or not focusing illusions and a failure to consider adaptation actually alter non-patient health state preference scores.

First, a study conducted by Ubel, Loewenstein, and Jepson found that reducing focusing illusions did not cause non-patients evaluations of quality of life to increase (2003). This group was careful to note that this study alone does not refute the potential influence of a focusing illusion, suggesting that the lack of impact could be due to the fact that the researchers did not adequately correct for focusing illusions (Ubel, Loewenstein, & Jepson, 2003). However, they also noted that they did not consider this outcome likely.

Another important problem is that both the focusing illusion might be impactful solely in cases where a health state has a damaging impact on one particular area of life, but leaves others unscathed. Hence the potential for a focusing illusion in an example like below-knee amputation where mobility would be greatly impacted, but one's ability to function mentally and socially would be unimpaired. Similarly, we can see potential problems with focusing illusions in health states like deafness, blindness, anxiety disorders, and other health states that particularly affect one area of life. A focusing illusion or an overweighting of physicality would be a less likely cause of the disparity between patients and non-patients when the health state in question impacts many areas of life. An example could be an accident that leaves an individual both mentally and physically handicapped to the degree that they are unable to care for themselves or live unassisted ever again.

Furthermore, Ubel, Loewenstein & Jepson note that perhaps a better application for a focusing illusion would be trivial scenarios with short-term consequences, but not chronic, deeply damaging health states in which the consequences are much broader and longer-lasting (2005). This could account for the presence of a focusing illusion in other studies that examined more trivial matters (e.g. students' ability to predict how upset they would be about receiving a bad grade), and the fact that it seems to not play a role in the way that patients evaluate health states.

We are left with adaptation and its potential role in the disability paradox. Several studies have been conducted to test for the degree to which considering adaptation might change general population evaluations of health states. There are several ways to bring

adaptation to the fore in a non-patients mind in order to understand the impact it has on health state preference scores.

Ubel, Loewenstein, & Jepson asked all their respondents to consider adaptation, and come up with examples of ways they had adapted to adverse circumstances in their pasts (2005). This led them to increase their quality weights for health states. However, in one of the three experiments, they felt that this actually evidenced a problem, which caused them to doubt the conclusion that underestimating adaptation caused patients to inaccurately evaluate health states. They used a previous experiment done by Schkade and Kahneman in 1998 to serve as the accurate values they were attempting to replicate, and found that asking participants to consider adaptation actually pushed their answers away from those found by Schkade and Kahneman. This led them to the conclusion that though they could cause non-patients' responses to rise, this did not inherently imply that they were becoming more accurate, and suggested that it might in fact mean the opposite (Ubel, Loewenstein, and Jepson, 2005).

Another group of researchers also hoped to evaluate the effect of underestimating adaptation and correcting for it on non-patient responses. Damschroder *et al.* gave "adaptation exercises" to a survey group before giving them a health state valuation survey to see if it altered their results and brought them closer to those of patients (Damschroder, Zikmund-Fisher, and Ubel, 2005).

Damschroder *et al.* gave a survey in which respondents considered a person tradeoff (PTO) elicitation (Damschroder, Zikmund-Fisher, and Ubel, 2005). Respondents chose between saving the lives of 100 previously healthy patients or saving 100 patients with pre-existing paraplegia. Respondents were also asked to choose the number of

patients with pre-existing paraplegia they would have to be able to save to equal the value of saving the 100 previously perfectly health patients. This helps determine an indifference point, or the relative value people assign to perfectly healthy people versus people with paraplegia (Damschröder, Zikmund-Fisher, and Ubel, 2005).

Damschröder *et al.* found that post adaptation exercises non-patients placed significantly higher value on the lives of patients with paraplegia relative to what they had valued paraplegia before the adaptation exercises (Damschröder, Zikmund-Fisher, and Ubel, 2005). This was particularly significant when they were asked to consider new onset paraplegia. People were far more inclined to save the lives of someone with existing paraplegia than someone with new onset of paraplegia (Damschröder, Zikmund-Fisher, and Ubel, 2005). This could suggest that increased understanding of adaptation could alter people's evaluations of health states. However, the warnings given by Ubel, Loewenstein, & Jepson about shifts in responses not necessarily representing an increase in accuracy of the answers still stand as an important caveat to consider (2003).

As with patients, it is difficult to say whether or not these potential errors truly lead non-patients to incorrectly evaluate the quality of life attached to various health states. A focusing illusion could mislead predictions of quality of life, but the evidence of its presence is hazy and thus it is hard to determine whether or not a focusing illusion alters non-patient responses. The evidence for non-patients underestimating the power of emotional adaptation is somewhat stronger, but still leaves us unsure as to whether or not it is a sufficiently confounding factor so as to consistently make non-patients significantly underestimate the quality of life experienced in health states. Thus it seems that neither

factor is entirely conclusive in determining whether or not non-patients evaluations are inaccurate reflections of the quality of life experienced with different health states.

Patients vs. Non-Patients

After considering the possible factors that could lead either group to over- or under-predict the quality weights of health states, we can see that both groups have potentially confounding factors influencing the way they respond (Ubel *et al* 2005, Ubel, Loewenstein & Jepson, 2003, Damschröder, Zikmund-Fisher, and Ubel, 2005). We can see potential areas for error in the responses of either group, and this is useful in that it allows us to see areas for increased accuracy. However it gives us little in terms of deciding which group's evaluations to heed. Many authors have posited reasons why one group is "wrong," but there are counterarguments to any point that might guide us to choose one side over the other. Even if we can eliminate confounding factors from the answers of either group, it does not seem that we can utterly eliminate the discrepancy in responses.

Perhaps neither group is wrong. We need one number for both clarity and for use in cost-effectiveness analysis, but that does not imply that one group's evaluations are inaccurate. In his essay "Ethical Issues in The Use of Cost-Effectiveness Analysis for the Prioritization of Health Care Resources" Dan Brock briefly mentions what he calls "the perspectives problem" (Brock, 2004). He suggests that because patients have adapted to their health states, they consider those health states from entirely different but equally

accurate perspectives as those of non-patients, who consider the same health state from the perspective of hoping never to experience it. Having a health state may have fundamentally changed a patient's preferences; they may be aware that they would have rated that same health state differently from the perspective of full health, but that does not change the way they now consider it (Brock 2004, Ubel, Richardson, and Pinto Prades, 1999).

Let's consider an example. Imagine a person with deafness. Before her hearing loss, she might have loved to go to concerts and play the piano. She derived a great deal of joy from these two activities, and without them, her life is different. Now that she is deaf, she has taken up painting, started teaching English sign language, and has joined a book club. She might miss playing the piano, but she is grateful for the fact that she isn't bothered by the road construction outside her apartment. Her evaluation of her quality of life valued concerts and playing the piano very highly before she lost her hearing. She would have imagined deafness as terrible since it would rob her of these two important parts of her life. Now she considers different activities and aspects of her life, and her evaluation of deafness comes from an entirely different perspective. Neither judgment is wrong, as both reflect different perceptions of deafness.

Like a patient, a non-patient looks at a health state and considers that health state relative to their state of full health. Their answers are two sides to the same coin. They may underestimate how well they will adapt to having that health state, but that does not make their desire to avoid that health state that much less intense.

Consider an analogy. Two people are standing at opposite corners of an intersection, waiting to cross the street. While they are waiting, there is an accident and

two cars collide in the middle of the intersection. When the police are writing the accident report, they ask both of the pedestrians for their account of the accident. Both tell the same basic story; two cars crashed into one another. But each person has a slightly different take on the accident; which car swerved more, which one braked first. This is because the pedestrians are quite literally looking at the accident from different angles. We might check to see if there are any reasons that one person might make a mistake in their version of the accident; perhaps one person is partially blind, or one of them looked down at their phone right at the instant of collision. But if we do not have clear evidence that one person is erroneous in their report, we are apt to assume that both accounts are reasonable. The same logic makes sense in the case of considering how a patient and a non-patient evaluate health states – they are looking at the situation from different angles.

If we decide that it is true that neither side is wrong; then we cannot denounce one group and choose the other on those grounds. Instead, we turn to considering the ethical and economic impacts of choosing one group versus the other. Let's consider first what it means to choose the non-patient responses.

Ethical and Economic Arguments for Choosing Patient or Non-Patient Evaluations

Quality weights for health states have traditionally been drawn from general population evaluations. As Gudex *et al.* put it, the fact that general population assessments of health state quality weights was a matter of “general agreement” (1996).

It was not until recently that the differentiation between patient and non-patient responses began to receive intense examination and researchers began to argue for the patient side.

On the non-patient side, some argue that non-patient assessments are not tainted by the same self-interest that might guide a patient to misrepresent their health state. However, this does not make sense in the context of the disability paradox. If patients were attempting to act in their own self-interest and garner more resources and spending towards their health state, it seems that they would rank them as worse than non-patients do. But as we know, this is the opposite of how the two respondent groups compare.

Another point is that non-patients are all potential patients for some health state, which gives them a right to help determine the level of resources allocated to any health state (Gudex *et al.* 1996). Furthermore, since cost-effectiveness analyses that rely on quality weights can and do play a role in government spending on healthcare, non-patients as taxpayers have the same right as patients in helping determine, albeit indirectly, how their tax dollars are divvied out (Nord, 1992, Gudex *et al.*, 1996).

Alternatively, patients provide informed, knowledge- and experience-based preferences, while non-patients can only provide predictions (Brock, 2004). Put this way, it seems obvious. In another scenario, we would easily choose the educated, informed response over the guess. However, there are reasons, such as the ones discussed prior, which lead people to believe that patients quality weights might be overestimates of their quality of life, and thus caution against assuming that the patient answers are inherently superior to those of non-patients.

First, let's consider the ethical and economic implications of choosing non-patient quality weights. Because their preference scores are lower, health states as evaluated using non-patient QALYs will be considered relatively more severe than they would be if we took patient evaluations. As was previously explained, this means that it will be more valuable to cure patients and return them to perfect health according to cost-effectiveness analysis. This means that we will most likely allocate more resources to spending and coverage of treatment costs because we consider the health states more severe.

However, many people have noted that this can actually pose a disadvantage to patients. Dan Brock, in his essay "Ethical Issues in The Use of Cost-Effectiveness Analysis," provides several cases in which a lower QALY score could disadvantage a patient (Brock, 2004).

First, a lower QALY weight will provide less benefit for treatments that extended life the same number of years (Brock, 2004). Imagine two people. Person A has severely impaired vision, giving her a QALY weight of .809 (Salomon *et al.* 2010). Person B is in perfect health, thus she has a QALY score of 1. If you were to give Person A five more years of life, she would gain 4.045 QALYs (ignoring discounting). If you were to give Person B five more years of life, she would gain 5 QALYs. Thus you should choose Person B because of her greater QALY gain. Both the people need the same treatment, but Person A is discriminated against because she already has a disability.

This seems to pose a moral problem in that we are placing less value on the life of a person with a disability purely because they have a disability. If we choose not to extend their lives thanks purely to their disability making it less cost-effective, then we seem to be considering that life less valuable, which some find morally problematic.

However, in his essay “Health Care Resource Prioritization and Rationing: Why Is It So Difficult?” Dan Brock notes that this worry about the undervaluing of life-extension for disabled people does not take into account the differences in the way we consider life-saving treatments and health-improving treatments (2007). If we truly do not differentiate between the two concepts: “...we should be indifferent between saving one person’s life who will live one year in full health, and preventing 20 people from suffering...hearing loss [with a QALY of .95]...for one year” (Brock 2007). Yet people tend to prefer saving the life over preventing the hearing loss, which seems to indicate that the two are not truly considered equivalent. This is called the rule of rescue, referring to the general tendency of people to favor life-saving interventions over health-improving treatments. Even if a life extension treatment is significantly more expensive, our psychological tendency holds fast; we will choose the life extension (Brock, 2007, Ubel, Richardson, and Pinto Prades, 1999)

Our largely steadfast adherence to the rule of rescue seems to undermine the worry that choosing lower quality weights will undervalue life-saving interventions. Unless we abandon our tendency to honor the rule of rescue and start strictly following cost-effectiveness analysis in our decisions regarding life-extension versus health improvement, we will not place a significant disadvantage on disabled persons.

In sum, choosing non-patient values means we will spend more overall on treatments for disabled persons. If we choose this option, we need to consider the economic trade-offs that will come with allocating more to individual health states. This money will have to come from somewhere, and whether that is from spending less on

health states we consider less severe or from other areas of the budget is something we will have to decide if we choose this option.

If non-patient assigned quality weights are indeed overestimating the negative impact of health states and underestimating the quality of life experienced in different health states, then we run the risk of overspending on health states that do not need as much as we are allocating to them, which could create inefficiencies in health spending and draw money away from other areas where it is more desperately needed.

Now let's turn to the other side, and consider the ethical and economic implications of choosing patient-assigned quality weights. If we choose patient responses, we are choosing the overall higher quality weights. This means that most health states will be considered less serious. The cost-effectiveness ratios will be less favorable, as less benefit will be returned from curing a health state with a higher quality weight. Thus we will allocate less spending to each health state.

Choosing patient evaluations of health states eliminates some of the concerns about disadvantaging patients; particular in terms of life-extension treatments as the disadvantage will not be as severe, relatively speaking. The disadvantage to patients thanks to cost-effectiveness analysis equating quality to value is still present and still needs to be addressed, but between the two options, choosing to heed patient responses will provide a better ratio.

It also needs to be considered that considering health states as less serious may present problems. Considering a health state as less serious will result in less funding partially due to an inferior cost-effectiveness ratio. This decrease in spending may preclude a patient from adapting in the ways patients were previously capable of doing,

particularly if those patients' adaptation was at all aided by resources that stemmed from that larger amount of spending.

So to sum up the results of choosing patient values, we will consider health states less serious and thus spend less to help treat them, though we may spend relatively more on life-extension treatments. If patients are overestimating their quality of life, this choice would mean underspending on health states that actually need more resources allocated towards them and overall more money.

Choosing between patient and non-patient evaluations of health states is enormously challenging task, and one that many authors have spent a great deal of time and energy debating, weighing in on where potential areas for over or underestimation are on either side of the equation. The question becomes even more complicated when we realize that the disability paradox is not universal. There is a notable exception to the disability paradox; depression defies the rule of patients placing their quality of life as higher than non-patients, with patients rating their quality of life as significantly lower than the ratings given by non-patients. In the next section we will discuss this anomaly and examine the data that demonstrates how depression defies the disability paradox.

CHAPTER III: THE ANOMALY OF DEPRESSION

We see the Disability Paradox almost everywhere – it seems that, almost without exception, there is a discrepancy between patient and non-patient responses. A notable exception to this is depression.⁴

Unlike most other health states, non-patients give relatively higher values to the quality of life an individual will experience when they have depression while patients give lower values. To put it roughly, someone without depression might say that depression is a 4 on a scale of 1 to 10, while someone with depression would say it is a 2 on the same scale.

A study done by Pyne *et al.* found evidence of significant differences in how patients versus non-patients scored depression in terms of the attached quality weight (2005). The researchers employed both Standard Gamble and Rating Scale questions to four different groups: general population respondents, patients who had experienced but were not currently experiencing depressive episodes, patients dealing with mild to moderate depression, and patients with moderate to severe depression. The results

⁴It seems worth exploring whether or not other mental illnesses produce the same discrepancy as depression. However few studies have been done comparing the quality weights of mental illnesses given by patients versus non-patients, and since some such studies have been done for depression, this paper will focus on depression.

showed that, particularly for patients with severe depression, there was a significant gap between the individuals and the patients (Pyne *et al.*, 2005).

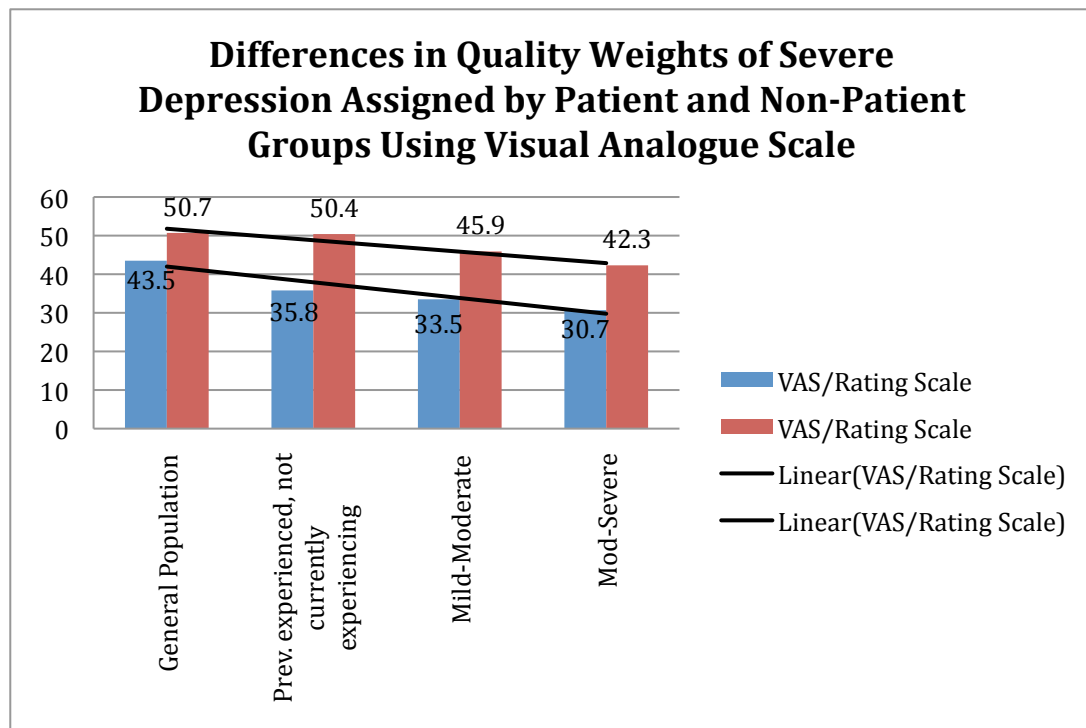
In order to create the hypothetical descriptions of the levels of depression that were given to the respondents, the researchers drew upon the Patient Health Questionnaire-9 (PHQ-9) and the Medical Outcomes Study SF-12. These questionnaires are both ways of measuring an individual's health status, and contain questions related to the symptoms of depression. The researchers used the symptoms in these questionnaires and the ways in which people at different levels of depression responded to create descriptions of the symptoms and facets of mild, moderate, and severe depression. This created a total of 6 different descriptions, with one description based on each questionnaire for each of the three levels of depression. It is interesting to note that the PHQ-9 descriptions tended to produce lower quality weights overall (Pyne *et al.*, 2005).

There is a clear overall trend downwards in the preference scores assigned to the three levels of depression evaluated. Furthermore, there is also a clear trend in lower preference scores given by respondents with a more severe degree of depression.

In the SF-12 health states, the responses given by the general population respondents and the moderate to severely depressed patients were significantly different across every level of depression that was described.⁵ This trend was true for both rating scale and standard gamble responses (Pyne *et al.*, 2005).

⁵ According to the χ^2 test done by Pyne *et al.* to compare the differences between the general population responses and the responses of the moderate to severely depressed respondent group, all three levels were significantly different. For mild depression, the responses were 89.5 versus 83.6, $p=.04$; for moderate depression 72.2 versus 62.7, $p=.001$; and lastly, for severe depression 50.7 versus 42.3, $p=.02$.

In the PHQ-9 health states, the rating scale responses held a general trend of higher preference scores for depression health states from the general population group compared to lower scores from the groups with current depression. Five of the six health states showed this trend. The standard gamble questions did not provide as clear of a trend line, however the average scores from the general population respondents for mild and moderate depression health states were significantly higher than the preference scores given by moderate to severely depressed respondents. The comparison between the preference scores for severe depression can be seen in the following graphs (Pyne *et al.*, 2005).



The above graph shows the how each of the four different respondent groups rated severe depression on a Visual Analogue Scale. As previously noted, the respondents considered two different descriptions of depression separately; the PHQ-9 is in red and the SF-12 is in blue. There is a clear downward trend in the responses; if they were

accurate to the Disability Paradox, the trend should go in the opposite direction. The table with all the original data and all the other graphs showing how respondents rated mild, moderate, and severe depression using Standard Gamble and also how the respondents rated mild and moderate depression using Visual Analogue Scale are in Appendix 1, Graphs 1.1-1.6.

While determining *why* we underestimate depression is interesting, it is not within the scope of this paper. Divining that particular “why” would require performing additional carefully constructed surveys (and thus require resources and time currently not at hand). Thus we are forced to set it aside for now, and answer another, equally important question; what are the implications of depression’s unusual preference scores?

As was previously discussed, the difference between patient and non-patient quality weights creates problems, particularly when we consider the distinctly different cost effectiveness analyses we will see when we use the different numbers. In the case of depression, the relatively anomalous nature of the differences in answers given by the two respondent groups makes this particularly important.

We previously discussed the ethical and economic implications of choosing either patients or non-patients and saw how those differentially affected spending. In most cases, our traditional tendency to go with the general population ratings will lead us to consider health states more severe than they actually are, and thus to consider it more cost-effective to treat them. But in the case of depression, we see the exact opposite occur. If we take the non-patient evaluation of depression, we run the risk of underestimating the severity of depression and thus spending less than we should.

The previous explanations of the implications of choosing either side, patients of non-patients, now seem odd because the results of choosing either side are now associated with the other, opposing argument. Our arguments for choosing patient values on the grounds that they are more efficient and will not run the risk of overspending stand in opposition to what happens in the case of depression. A similar problem appears when we consider our arguments for non-patient responses.

This leads me to wonder if choosing between different quality weights on the basis of which group of respondents provided them could be the wrong method of choosing. We already discussed the implications of choosing one group or the other, but we should consider that question more abstractly. If instead of choosing between patients or non-patients we consider choosing between a higher quality weight and a lower quality weight, we can see similar implications but without the confusing arguments tied to patient or non-patient responses.

CHAPTER IV: REFRAMING THE QUESTION

Higher versus Lower Quality Weights

If we reframe the question of which quality weight to choose as a decision between a higher bracket and a lower bracket as opposed to deciding between the groups that defined those, the way we consider the choice becomes slightly different. Instead of examining both patients and non-patients and trying to find reasons that either group would overestimate or underestimate health related quality of life, we consider primarily the ethical and economic impacts of choosing a higher versus a lower quality weight to assign to a health state.

To briefly recap, choosing a higher quality weight implies that health states are relatively less serious and have less of an impact on quality of life. Thus treating them is less appealing as there is a smaller cost-effectiveness ratio. However, higher quality weights also make life-extension treatments relatively more appealing as there is a larger QALY gain. Alternatively, lower quality of life values imply that health states should be taken more seriously and more spending should and will be allocated to address them, particularly health states on the lower end of the quality of life spectrum, because curing them produces a larger, more appealing cost-effectiveness ratio.

This could cause a scenario where, from a numerical basis, life-extension treatments will be less valuable as there will be a smaller QALY gain from saving the life of a patient who has a health state with a lower quality weight. However, as previously

noted, the mathematical explanation does not fit with how we actually treat life-extension treatments, and thus this worry is largely unfounded.

Thus, the question really comes down to whether we want to spend more or less on people with disabilities. There are downsides to either side, largely hinging on our uncertainty about the accuracy of either set of quality weights. If we choose the lower QALY weights and they are inaccurate, we run the risk of overspending on health states that are not as serious as we are considering them. If we choose the higher quality weights, we run the risk of underspending on health states that are more serious than we say.

As we have already seen, the choice is not an easy one. There are benefits and drawbacks to either side. The evidence is split between the two, leaving us with no empirical foundation to fall back on. Given that there are few differentiating factors that would tip the scale to either side, it seems that our moral imperative is to choose the option that is “better safe than sorry” and go with the lower quality weights.

The risk of underspending on health states is worse, morally speaking, than the risk of overspending on health states. If in reality the higher evaluations of health states are inaccurate and the actual quality of life one experiences is lower, we would have allocated inadequate resources towards handling a health state that needed more resources, resulting in a failure to serve those who need it. Alternatively, if we take the lower quality weights, we might give more resources towards treatment and research related to health states that don't need as much help as we are giving them. While this is inefficient, it is the morally responsible option as it does not run the risk of leaving someone in need without the resources they require.

Part of the rationale for this comes from the way in which depression defies the disability paradox. The fact that depression does not fit into the mold that holds most other health states seems to imply that there is something significantly different and unusual about this case. If we used the higher quality weight to evaluate depression, we would not consider it as serious as its anomalous nature seems to suggest it might be. If we have reason to believe that depression even *could* be worse than the general population thinks, that seems to suggest that we need to heed that and devote more resources and spending to addressing it.

The argument is this; given that, thanks the disability paradox, we need to choose between two quality weights for health states, one indicating higher quality of life and one indicating lower. If there is a chance that the lower quality of life accurately reflects the quality of life one experiences with a health state, we need to assume that all patients who experience that health state might experience that low quality of life, and thus we need to be able to provide resources and treatments to accommodate such a level.

Consider an analogy. A friend recently paid for your meal when you forgot your wallet. She doesn't have the receipt but you both know that the amount you owe her is around \$10. You have \$29 in cash when you see her again - \$9 in ones or a \$20. She has \$9 on her person. This means that you can pay her either \$9 or \$11. In general, we consider it better to go with \$11. Because we don't know what the exact amount you owe her should be, we think it's better to give her slightly more than might be owed, rather than run the risk of underpaying her.

The parallels for this example are fairly clear. First, in a situation where we do not know what exact amount is needed, we find it more morally acceptable to potentially

overestimate instead of underestimating. Second, the fact that we are determining how much to pay back a friend illustrates an important facet. If the person were paying us back for the meal instead of vice versa, we might find it acceptable to accept the 9 dollars; we are more willing to incur that risk on ourselves. In general, we tend to find it okay to accept individual risks, but not reasonable to assign those same risks to others.

If we have this intuition in a case that does not seem to hold great moral weight – the difference of a dollar is arguably trivial – it seems reasonable to extrapolate this same reasoning to scenarios where that potential extra dollar represents a great deal more. We tend to agree that health is important, whether we think it is valuable as a means to an end or merely in itself. Given this, it seems there is greater moral urgency in helping people have and maintain full health. So even if we need to overspend slightly to help them achieve that, it seems to be the morally sound decision to make, and we should err on the side of overspending in these cases as opposed to under-spending.

CHAPTER V: ECONOMIC TRADE-OFFS

This decision leaves an enormous problem looming. If we choose to “overspend,” the money must come from somewhere. We must consider where the trade-off will lie when we choose to heed the lower quality weights assigned to health states. We could either draw the money from less severe health states, thus rationing our limited health care resources, or we could find others areas in the budget from which to draw it. First, we consider rationing, and whether it is a viable option.

Choosing to value health states at a lower quality of life will, as previously explained, guide us towards allocating more resources towards treating them, and this might result in withdrawing provisions of beneficial care for health states evaluated as having higher quality of life. Prioritization of health care already tends to result in rationing, but this might exacerbate that and make the rationing process harsher against providing healthcare resources to people in health states that are associated with higher quality of life.

Rationing in health care refers to the process of allocating and distributing limited resources in healthcare. The trade-off mentioned above may involve choosing to allocate more healthcare resources towards more serious health states, and potentially withdrawing healthcare resources from health states that we consider less serious, and which have higher quality weights. Rationing has a bad reputation, but as Dan Brock points out in his essay “Health Care Resource Prioritization and Rationing: Why Is It So Difficult?” this reputation seems to be less due to a legitimate anger at the concept and more thanks to a lack of accurate understanding of both what rationing is and its ubiquity (2007). Brock also notes that people’s fear of rationing is driven by a worry that if they

support rationing they will someday be denied care they want or need, rather than a real belief that there is a more ethical way to distribute limited resources (2007).

Rationing occurs constantly, both in day-to-day life and in health care. When we choose to buy one food at the grocery store over another, we are rationing our limited budget for groceries between different foods. In health care, Brock offers the example of when doctors decide whether or not to offer a patient an MRI in order to determine if their migraines are the sign of a brain tumor (2007). If they determine that the likelihood of a tumor is too low to merit the expense of an MRI, they are effectively rationing treatments.

Rationing is a necessary means of allocating limited resources. Given that our health care resources are limited by our government's budget, we must use rationing in order to allocate resources efficiently and ethically – realistically, we seem unable to find another method capable of operating within the boundary of limited resources that would produce more ethical and efficient resource distribution. While on the surface it may seem bad to pull resources from one health state to help another, it seems like an inevitable facet of a system with limited resources.

If we cannot find a way a morally acceptable way to use money that is already allocated to health care spending, we still have a moral imperative to find the money to treat these health states according to their lower quality weight. This may mean we need to draw money to health care from somewhere else in the federal budget.

At first this seems tricky, but closer examination allows us to see where we should draw the money from when re-allocating funds from elsewhere in the budget. As previously noted, we seem to have a significant moral imperative to help people have

access to and maintain health. When we consider it carefully, we can see that we spend on many things that do not hold the same moral imperative, such as public parades or even public parks. While these things are valuable, they should come second to helping people have access to different treatments for health states. While I do not have empirical evidence to guarantee that this is where additional funding for research would come, it seems likely that if we chose to spend more on treatment and research for health, we would draw from areas of the budget that fall lower on the scale of moral importance. It is worth remembering, at this point, that although we briefly considered the concern that lower quality weights for different health states would place less moral urgency on extending the life of patients with those health states, we realized that the rule of rescue refutes this. Though cost-effectiveness analysis alone might tell us to draw money for health-improvement treatments from life-extension treatments, this is an unlikely decision for health economists to make.

It is also plausible to consider that allocating more money to health states that are considered more severe may be costly in the short-term but provide long-term benefits that will ultimately benefit healthcare as a whole. Perhaps the additional resources allocated to addressing depression will result in the development of a new, most cost-effective treatment that will ultimately require fewer resources to treat depression in a more effective way than ever before. If this hypothetical new treatment required the additional resources to develop but ultimately freed up resources to go to other health states, it seems like the initial expense was worth it. This is not a guaranteed outcome, but one worth exploring as a possibility.

We should also consider carefully what health states would receive more money. We have traditionally taken the general population preference scores for health states, which means that in choosing to honor the lower quality weights we would not be changing our evaluations, in most cases (Gudex *et al.* 2006, Sullivan 2003, Boyd *et al.* 1990). The case that would stand out as receiving more funding would be depression. Depression would be one of the few health states that we would switch to considering as having a significantly lower preference score and fewer QALYs. Thus it seems possible that while we would have to increase funding slightly, it would not be a huge jump.

CONCLUSION

Good health is important in such a deep, universally desired way that statements such as “if you’ve got your health, you have everything” have been created to express its importance. Since good and bad health states occur with some degree of randomness, we cope with this by pooling our resources through private or governmental insurance plans. While the desire for perfect health states is nearly infinite, the resources available to achieve those desires are limited, even when risks are pooled and governments are involved.

We are left with a difficult problem of how to maximize health results with those limited funds. One of the best methods for sorting and allocating resources involves assessing the impact of various health problems on the quality of life. This results in QALY scores, which promise to help us work towards maximizing the quality and duration of results for health spending.

We employ QALYs gained as a measure of the cost effectiveness of various treatment choices. Obviously enough, rationing decisions are fraught with emotion. The QALY approach is a method that attempts to rationally sort through various health care choices and find the one that makes most economic sense.

As I have noted here, there are a number of pitfalls that can arise within the QALY system. Chapter II focused on a particularly profound issue, the Disability Paradox. This is the fact that patient and non-patient groups provide distinctly different quality weights to health conditions as imagined versus health conditions as actually experienced. Thus the paradox, people actually suffering almost any condition don’t report it to lower the quality of their life as much as others would assume. Naturally, this

has sparked a great deal of debate over which group's quality of life numbers should be used and whether there are psychological factors in either the patient or observer group that make their assessments incorrect.

I believe that we should not broadly consider either patients or non-patients to be wrong. This position is supported, in part, by the evidence surrounding patient assessment of depression, which defies the disability paradox. Unlike almost all other health states, depression is routinely reported to be a much worse health state by its sufferers than how it would be assessed by non-patients. Therefore, whichever QALY set (patient or non-patient) was chosen for all negative health states would produce an opposite result in spending for depression than it would for all other maladies.

This has caused me to wonder if it is not better to choose from a more general viewpoint. Instead of trying to choose between the patient and non-patient QALYs, perhaps we should consider a blended approach, having a third party choose the higher or lower weights. We should consider the economic and moral costs of defaulting to the lower number of the two possibilities.

Let's briefly consider the implications defaulting to the lower number would have. QALY systems have traditionally used the general population's quality weights for health states, meaning we already use a direr, lower numeric assessment for most health states. Since we already largely default to the lower number, the current system would, in most cases, err on the side of over-spending. The salient exception is depression, where choosing the patients' lower number would call for more spending.

If we make this choice we face some important decisions. The extra funding to cover any additional treatments that we would choose to fund has to come from somewhere. Whether it would come from funds pulled from health states we consider less dire or from other areas of a federal budget we consider less morally urgent is unknown. The moral imperative to act by allocating more resources to those suffering from depression is clear. What the source of those funds would be is less clear.

The fact that depression defies the disability paradox suggests that the malady has unique characteristics which have gone unrecognized. It's worth noting that depression not only gets a lower quality rating from patients, it gets one of the lowest quality ratings of any affliction rated by either group. Our general adherence to the rule of rescue applies, since these patients are among the worst off, their needs have a strong moral claim on receiving more resources.

BIBLIOGRAPHY

Albrecht, Gary L., and Patrick J. Devlieger. 1999. "The Disability Paradox: High Quality of Life Against All Odds." *Social Science & Medicine* 48 (8): 977-988.

- Albrecht, Gary L., and Paul C. Higgins. 1977. "Rehabilitation Success: The Interrelationships of Multiple Criteria." *Journal of Health and Social Behavior* 18 (1): 36.
- Asarnow, J. R. 2005. "Effectiveness of a Quality Improvement Intervention for Adolescent Depression in Primary Care Clinics: A Randomized Controlled Trial." *JAMA: The Journal of the American Medical Association* 293 (3): 311-319.
- Badia, X. 1999. "Feasibility, Validity and Test-Retest Reliability of Scaling Methods for Health States: The Visual Analogue Scale and the Time Trade-Off." *Quality of Life Research* 8 (4): 303-310.
- Badia, X., A. Díaz-Prieto, M. Rué, and D. L. Patrick. 1996. "Measuring Health and Health State Preferences Among Critically Ill Patients." *Intensive Care Medicine* 22 (12): 1379-1384.
- Barney, Lisa J., Kathleen M. Griffiths, Anthony F. Jorm, and Helen Christensen. 2006. "Stigma About Depression and Its Impact on Help-seeking Intentions." *Australian and New Zealand Journal of Psychiatry* 40 (1): 51-54.
- Baron, Jonathan, David A. Asch, Angela Fagerlin, Christopher Jepson, George Loewenstein, Jason Riis, Margaret G. Stineman, and Peter A. Ubel. 2003. "Effect of Assessment Method on the Discrepancy Between Judgments of Health Disorders People Have and Do Not Have: A Web Study." *Medical Decision Making* 23 (5): 422-434. doi:10.1177/0272989X03257277.

- Bleichrodt, Han, and Magnus Johannesson. 1997. "Standard Gamble, Time Trade-off and Rating Scale: Experimental Results on the Ranking Properties of QALYs." *Journal of Health Economics* 16 (2): 155-175.
- Boyd, N. F., H. J. Sutherland, K. Z. Heasman, D. L. Tritchler, and B. J. Cummings. 1990. "Whose Utilities for Decision Analysis?." *Medical Decision Making* 10 (1): 58-67.
- Brazier J, M. Deverill, C. Green, R. Harper, A Booth. 1999. A review of the use of health status measures in economic evaluation. *Health Technol Assess*: 3(9).
- Brock, Dan. 2003. "Ethical Issues in The Use of Cost-Effectiveness Analysis for The Prioritization of Health Care Resources." Essay. In *Making Choices in Health: WHO Guide to Cost-effectiveness Analysis*, 289-311. Geneva: World Health Organization.
- Brock, Dan W. 2001. "Two Moral Issues About Disability." *American Journal of Bioethics* 1 (3): 1-2. doi:10.1162/15265160152648125.
- Brock, Dan W. 2007. "Health Care Resource Prioritization and Rationing: Why Is It So Difficult?." *Social Research* 74 (1, Difficult Choices): 125-148.
<http://www.jstor.org/stable/10.2307/40971892?ref=no-x-route:4c6c20275a79dd64a8d61ef3ecf19382>.
- Brock, Dan W. 2009. "Cost-Effectiveness And Disability Discrimination." *Economics and Philosophy* 25 (01): 27. doi:10.1017/S0266267108002265.
- Broome, John. 1993. "Qalys." *Journal of Public Economics* 50 (2): 149-167.

- Burström, Kristina, Magnus Johannesson, and Finn Diderichsen. 2006. "A Comparison of Individual and Social Time Trade-off Values for Health States in the General Population." *Health Policy* 76 (3): 359-370.
- Damschroder, Laura J., Brian J. Zikmund-Fisher, and Peter A. Ubel. 2005. "The Impact of Considering Adaptation in Health State Valuation." *Social Science & Medicine* 61 (2): 267-277. doi:10.1016/j.socscimed.2004.11.060.
- Doesschate, Mascha C. Ten, Maarten W.j. Koeter, Claudi L.h. Bockting, and Aart H. Schene. 2010. "Health Related Quality of Life in Recurrent Depression: A Comparison With a General Population Sample." *Journal of Affective Disorders* 120 (1-3): 126-132.
- Dolan, Paul, Rebecca Shaw, Aki Tsuchiya, and Alan Williams. 2005. "QALY Maximisation and People's Preferences: a Methodological Review of the Literature." *Health Economics* 14 (2): 197-208.
- Drummond, Michael. 1997. "Cost-utility Analysis." Essay. In *Methods for the Economic Evaluation of Health Care Programmes*, 3th ed. 137-179. Oxford: Oxford University Press.
- Essink-Bot, Marie-Louise, Gouke J. Bonsel, and Paul J. Van Der Maas. 1990. "Valuation of Health States by the General Public: Feasibility of a Standardized Measurement Procedure." *Social Science & Medicine* 31 (11): 1201-1206.
- Fellinghauer, Bernd, Jan D. Reinhardt, Gerold Stucki, and Jerome Bickenbach. 2012. "Explaining the Disability Paradox: a Cross-sectional Analysis of the Swiss General Population." *BMC Public Health* 12 (1): 655.

- Froberg, Debra G., and Robert L. Kane. 1989. "Methodology for Measuring Health-state Preferences—III: Population and Context Effects." *Journal of Clinical Epidemiology* 42 (6): 585-592.
- Gonzalez, H. M., W. A. Vega, D. R. Williams, W. Tarraf, B. T. West, and H. W. Neighbors. 2010. "Depression Care in the United States: Too Little for Too Few." *Archives of General Psychiatry* 67 (1): 37-46.
- Green, C., J. Brazier, and M. deverill. 2000. "Valuing Health-Related Quality of Life: A Review of Health State Valuation Techniques." *PharmacoEconomics* 17 (2): 151-165.
- Gudex, C., P. Dolan, P. Kind, and A. Williams. 1996. "Health State Valuations From the General Public Using the Visual Analogue Scale." *Quality of Life Research* 5 (6): 521-531.
- Guyatt, Gordon H. 1993. "Measuring Health-Related Quality of Life." *Annals of Internal Medicine* 118 (8): 622.
- Hadorn, D. C. 1991. "Setting Health Care Priorities in Oregon Cost-effectiveness Meets the Rule of Rescue." *JAMA: The Journal of the American Medical Association* 265 (17): 2218-2225.
- Kaplan, Robert M. 2005. "The Minimally Clinically Important Difference in Generic Utility-Based Measures." *COPD: Journal of Chronic Obstructive Pulmonary Disease* 2 (1): 91-97.
- Kaplan, Robert M., and James W. Bush. 1982. "Health-related Quality of Life Measurement for Evaluation Research and Policy Analysis." *Health Psychology* 1 (1): 61-80.

- Kopec, J. 2003. "A Comparative Review of Four Preference-weighted Measures of Health-related Quality of Life." *Journal of Clinical Epidemiology* 56 (4): 317-325.
- Krabbe, Paul F.m., Marie-Louise Essink-Bot, and Gouke J. Bonsel. 1997. "The Comparability and Reliability of Five Health-state Valuation Methods." *Social Science & Medicine* 45 (11): 1641-1652.
- Lenert, Leslie A., Jonathan R. Treadwell, and Carolyn E. Schwartz. 1999. "Associations Between Health Status and Utilities Implications for Policy." *Medical Care* 37 (5): 479-489.
- Link, B.G., J.C. Phelan, M. Bresnahan, A. Stueve, and B.A. Pescosolido. 1999. "Public Conceptions of Mental Illness: Labels, Causes, Dangerousness, and Social Distance." *American Journal of Public Health* 89 (9): 1328-1333.
- Luppa, Melanie, Sven Heinrich, Matthias C. Angermeyer, Hans-Helmut König, and Steffis G. Riedel-Heller. 2007. "Cost-of-illness Studies of Depression." *Journal of Affective Disorders* 98 (1-2): 29-43.
- Martin, Andrew J., P. P. Glasziou, R. J. Simes, and T. Lumley. 2000. "A Comparison Of Standard Gamble, Time Trade-Off, And Adjusted Time Trade-Off Scores." *International Journal of Technology Assessment in Health Care* 16 (1): 137-147.
- Morimoto, Takeshi, and Tsuguya Fukui. 2002. "Utilities Measured by Rating Scale, Time Trade-off, and Standard Gamble: Review and Reference for Health Care Professionals." *Journal of Epidemiology* 12 (2): 160-178.

- Nord, Erik. 1992. "Methods for Quality Adjustment of Life Years." *Social Science & Medicine* 34 (5): 559-569.
- Patrick, D. L., H. E. Starks, K. C. Cain, R. F. Uhlmann, and R. A. Pearlman. 1994. "Measuring Preferences for Health States Worse Than Death." *Medical Decision Making* 14 (1): 9-18.
- Perry, B. L., B. A. Pescosolido, J. K. Martin, J. D. Mcleod, and P. S. Jensen. 2007. "Comparison of Public Attributions, Attitudes, and Stigma in Regard to Depression Among Children and Adults." *Psychiatric Services* 58 (5): 632-635.
- Pyne, Jeffrey M., John C. Fortney, Shanti Tripathi, David Feeny, Peter Ubel, and John Brazier. 2009. "How Bad Is Depression? Preference Score Estimates From Depressed Patients and the General Population." *Health Services Research* 44 (4): 1406-1423.
- Revicki, D. 1998. "Patient-assigned Health State Utilities for Depression-related Outcomes: Differences by Depression Severity and Antidepressant Medications." *Journal of Affective Disorders* 48 (1): 25-36.
- Riis, Jason, George Loewenstein, Jonathan Baron, Christopher Jepson, Angela Fagerlin, and Peter A. Ubel. 2005. "Ignorance of Hedonic Adaptation to Hemodialysis: A Study Using Ecological Momentary Assessment." *Journal of Experimental Psychology: General* 134 (1): 3-9. doi:10.1037/0096-3445.134.1.3.
- Sackett, David L., and George W. Torrance. 1978. "The Utility of Different Health States as Perceived by the General Public." *Journal of Chronic Diseases* 31 (11): 697-704.

- Salomon, Joshua A., Theo Vos, Daniel R. Hogan, Michael Gagnon, Mohsen Naghavi, Ali Mokdad, Nazma Begum, et al. 2012. "Common Values in Assessing Health Outcomes From Disease and Injury: Disability Weights Measurement Study for the Global Burden of Disease Study 2010." *The Lancet* 380 (9859): 2129-2143.
- Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9, 340 –346.
- Schoenbaum, M. 2001. "Cost-effectiveness of Practice-Initiated Quality Improvement for Depression: Results of a Randomized Controlled Trial." *JAMA: The Journal of the American Medical Association* 286 (11): 1325-1330.
- Schwappach, David L. B. 2002. "Resource Allocation, Social Values and the QALY: a Review of the Debate and Empirical Evidence." *Health Expectations* 5 (3): 210-222.
- Schwartz, Carolyn E., Rita Bode, Nicholas Repucci, Janine Becker, Mirjam A. G. Sprangers, and Peter M. Fayers. 2006. "The Clinical Significance of Adaptation to Changing Health: A Meta-analysis of Response Shift." *Quality of Life Research* 15 (9): 1533-1550.
- Sherbourne, Cathy Donald, Jürgen Unützer, Michael Schoenbaum, Naihua Duan, Leslie A. Lenert, Roland Sturm, and Kenneth B. Wells. 2001. "Can Utility-Weighted Health-Related Quality-of-Life Estimates Capture Health Effects of Quality Improvement for Depression?." *Medical Care* 39 (11): 1246-1259.

- Stein, Mark S. 2001. "Utilitarianism and the Disabled: Distribution of Life." *Social Theory and Practice* 27 (4, Special Issue: Embodied Values: Philosophy and Disabilities): 561-578.
- Sullivan, Mark. 2003. "The New Subjective Medicine: Taking the Patient's Point of View on Health Care and Health." *Social Science & Medicine* 56 (7): 1595-1604.
- Turner, Nicholas, John Campbell, Tim J. Peters, Nicola Wiles, and Sandra Hollinghurst. 2013. "A Comparison of Four Different Approaches to Measuring Health Utility in Depressed Patients." *Health and Quality of Life Outcomes* 11 (1): 81.
- Ubel, Peter A. 1999. "LIFE-SAVING TREATMENTS AND DISABILITIES (Are All QALYs Created Equal?)." *International Journal of Technology Assessment in Health Care* 15 (4): 738-748.
- Ubel, Peter A., George Loewenstein, and Christopher Jepson. 2003. "Disability and Sunshine: Can Hedonic Predictions Be Improved by Drawing Attention to Focusing Illusions or Emotional Adaptation?." *Journal of Experimental Psychology: Applied* 11 (2): 111-123.
- Ubel, P. A., George Loewenstein, J. Hershey, J. Baron, T. Mohr, D. A. Asch, et al. 2001. "Do nonpatients underestimate the quality of life associated with chronic health conditions because of a focusing illusion?." *Medical Decision Making*, 21: 190 –199.
- Ubel, Peter A., George Loewenstein, Norbert Schwarz, and Dylan Smith. 2005. "Misimagining the Unimaginable: The Disability Paradox and Health Care Decision Making." *Health Psychology* 24 (4, Suppl): S57-S62. doi:10.1037/0278-6133.24.4.S57.

Untzer, Jrgen, Donald L. Patrick, Paula Diehr, Greg Simon, David Grembowski, and

Wayne Katon. 2000. "Quality Adjusted Life Years in Older Adults With Depressive Symptoms and Chronic Medical Disorders." *International Psychogeriatrics* 12 (1): 15-33.

Weinstein, Milton C., George Torrance, and Alistair Mcguire. 2009. "QALYs: The Basics." *Value in Health* 12: S5-S9.

Wit, G. Ardine De, Jan J.v. Busschbach, and Frank Th. De Charro. 2000. "Sensitivity and Perspective in the Valuation of Health Status: Whose Values Count?." *Health Economics* 9 (2): 109-126.

Zimovetz, Evelina A., Sorrel E. Wolowacz, Peter M. Classi, and Julie Birt. 2012.

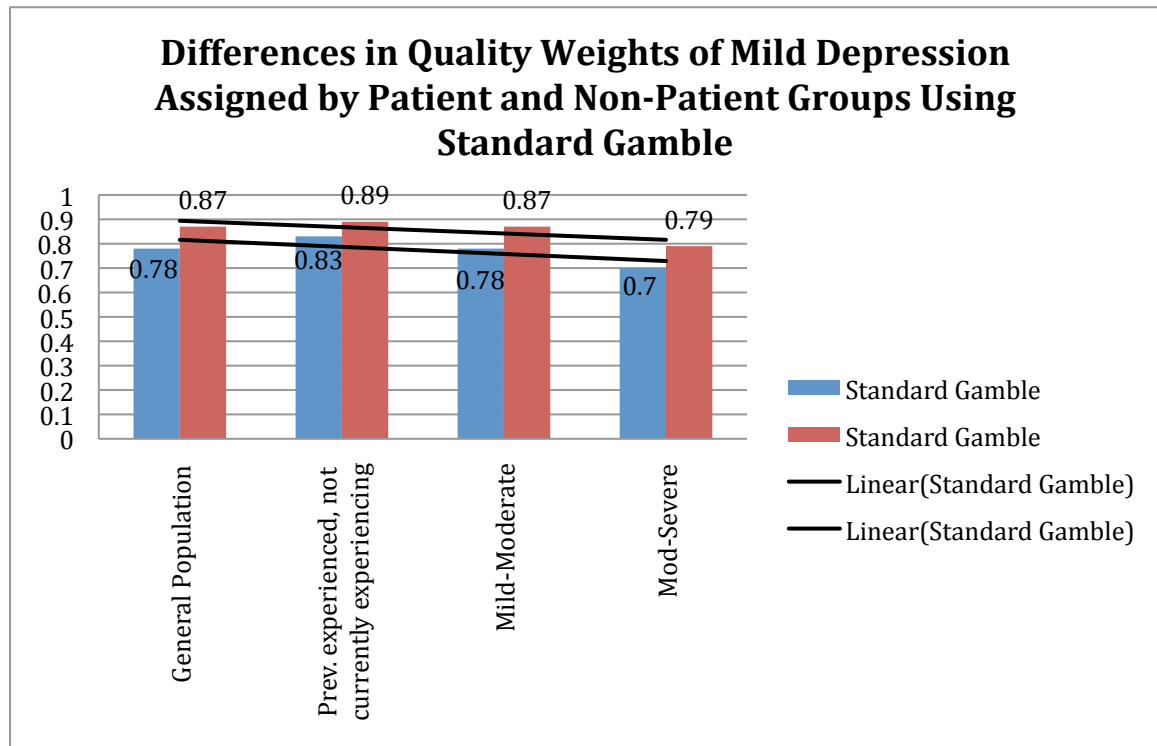
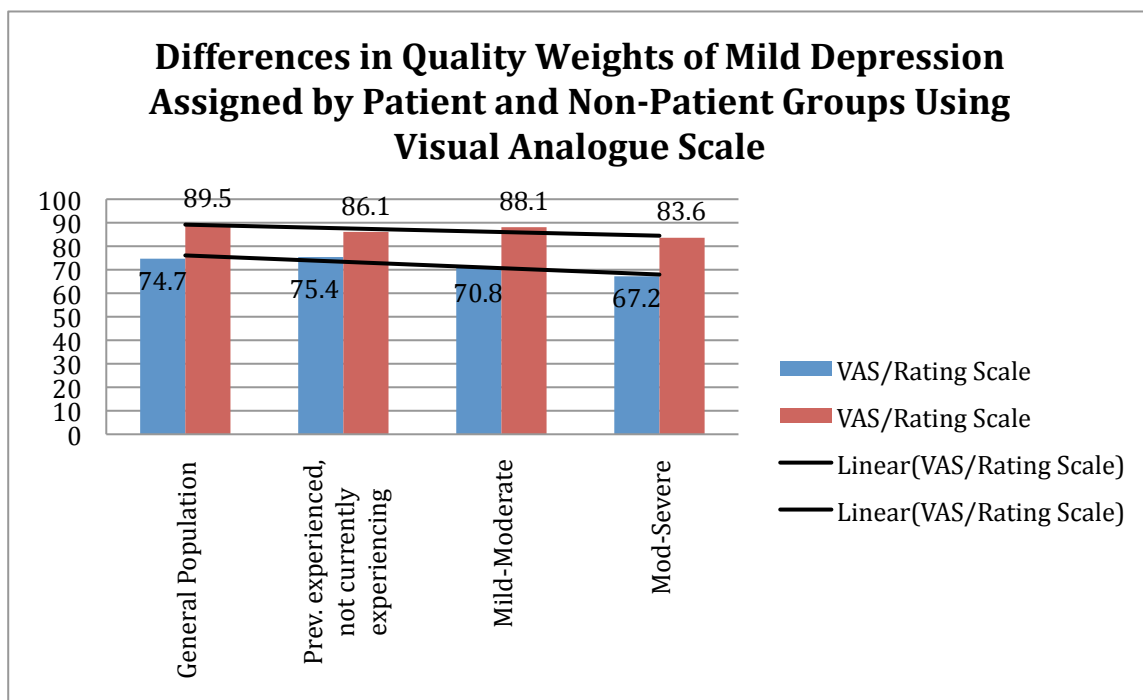
"Methodologies Used in Cost-effectiveness Models for Evaluating Treatments in Major Depressive Disorder: a Systematic Review." *Cost Effectiveness and Resource Allocation* 10 (1): 1.

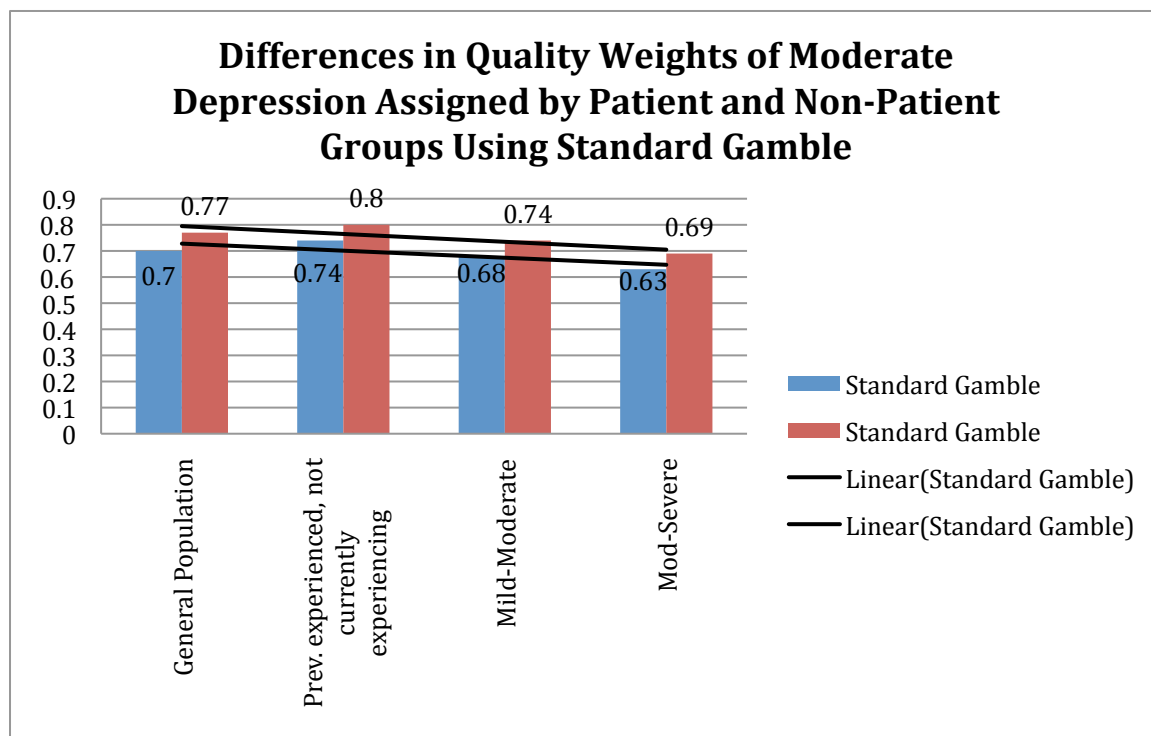
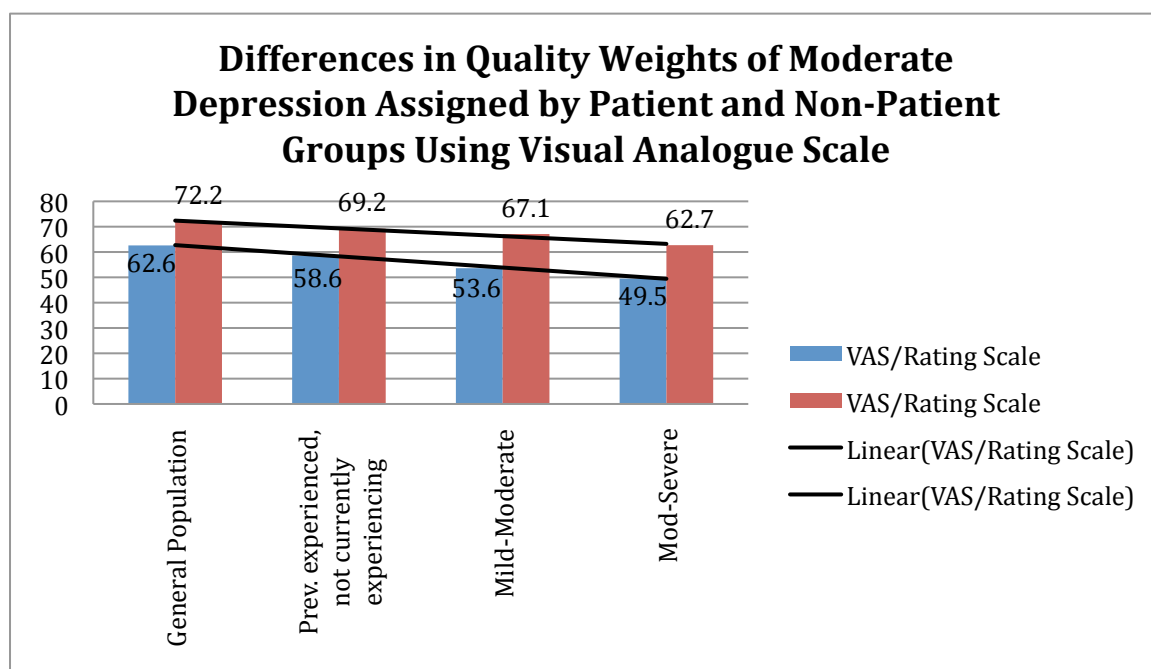
Appendix 1

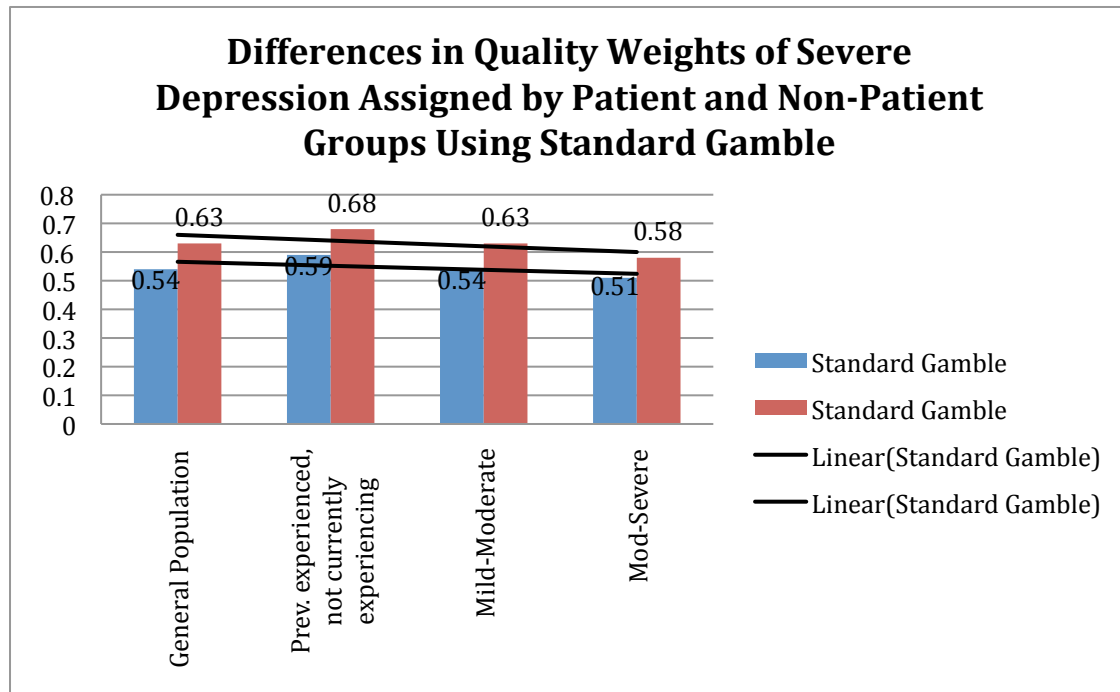
All data provided by Pyne *et al*, ““How Bad Is Depression? Preference Score Estimates From Depressed Patients and the General Population.”(2005). Data was collected and consolidated from their study.

Table 1.1: all VAS and TTO data from Pyne *et al*. 2005.

		Respondent Groups				
		Non-patients	Patients (have experienced or are experiencing depression)			
Preference elicitation question	Degree of Depression being rated	General Population	Prev. experienced, not currently experiencing	Mild-Moderate	Mod-Severe	Determiner of Degree of Depression in Respondents
Standard Gamble	Mild Depression	0.78	0.83	0.78	0.7	PHQ-9
Standard Gamble	Moderate Depression	0.7	0.74	0.68	0.63	PHQ-9
Standard Gamble	Severe Depression	0.54	0.59	0.54	0.51	PHQ-9
Standard Gamble	Mild Depression	0.87	0.89	0.87	0.79	SF-12
Standard Gamble	Moderate Depression	0.77	0.8	0.74	0.69	SF-12
Standard Gamble	Severe Depression	0.63	0.68	0.63	0.58	SF-12
VAS/Rating Scale	Mild Depression	74.7	75.4	70.8	67.2	PHQ-9
VAS/Rating Scale	Moderate Depression	62.6	58.6	53.6	49.5	PHQ-9
VAS/Rating Scale	Severe Depression	43.5	35.8	33.5	30.7	PHQ-9
VAS/Rating Scale	Mild Depression	89.5	86.1	88.1	83.6	SF-12
VAS/Rating Scale	Moderate Depression	72.2	69.2	67.1	62.7	SF-12
VAS/Rating Scale	Severe Depression	50.7	50.4	45.9	42.3	SF-12

Graph 1.1: Respondent valuations of mild depression via standard gamble**Graph 1.2: Respondent valuations of mild depression via visual analogue scale**

Graph 1.3: Respondent valuations of moderate depression via standard gamble**Graph 1.4: Respondent valuations of moderate depression via visual analogue scale**

Graph 1.5: Respondent valuations of severe depression via standard gamble**Graph 1.6: Respondent valuations of severe depression via visual analogue scale**